A Principled Call to Ban All Agentic AI from US Government (Federal, State, Local)

Christopher Hunt Robertson, M.Ed.

(Written with the Support of Advanced AI – ChatGPT and Perplexity)

In a recent essay (Chapter 10), I introduced a proposal for *Civic Digital Visibility First* - a principle for designing AI systems that help democracies see clearly without taking part in decision-making. That essay focused on the solution: a form of non-agentic Civic AI that illuminates conditions while leaving judgment where it belongs - with human institutions and citizens. But a solution only makes sense if the problem is clearly understood. The two chapters that follow - drawn from the e-book *Civic AI Visibility Infrastructure (Volumes I and II)* - attempt to define that problem in structural terms. (This is not a ban on automation or computation. It is a boundary on delegated judgment in domains of civic authority.)

Chapter 8 below examines a question that is increasingly treated as common sense: if AI can optimize logistics, reduce delays, and manage complexity, why shouldn't it help run government? The answer offered here is not that such systems are poorly designed or insufficiently regulated. It is that, when agentic AI operates near public authority - deciding who qualifies, who is prioritized, who is watched, or how power is applied - it crosses what this work calls the *civic ceiling*. Beyond that boundary, even well-intentioned systems do not merely assist governance. Over time, they begin to reconfigure it. This shift does not require failure or misuse. It follows a recognizable pattern: systems that assist begin to recommend; systems that recommend begin to be followed; systems that are followed become embedded; and systems that become embedded begin to govern. Chapter 8 maps how this progression - structural drift - appears across six emerging families of agentic systems in American governance, from eligibility engines to optimization tools to autonomous service agents. The claim is direct: these systems are not simply risky. By structure, they are incompatible with democratic admissibility as Civic AI.

Chapter 9 below addresses what might otherwise appear to be a gap in that argument. Even if governments themselves proceed cautiously, what happens when agentic systems are deployed by contractors, platforms, or adjacent institutions operating at the edge of governance? The answer proposed here is not to extend AI authority further, but to define a different role entirely. Non-agentic Civic AI, properly bounded, does not intervene, optimize, or co-govern. It watches. It observes patterns of system behavior, detects when they approach domains of civic judgment, and issues non-binding alerts to human authorities. Its function is not control, but visibility - maintaining the boundary between delegation in matters of function and human judgment in matters of rights, access, and power. Taken together, these chapters are not an argument against artificial intelligence. They are an argument about where certain kinds of AI do not belong near governance within a constitutional democracy.

If Chapter 10 offered a proposal for how Civic AI might serve as a form of public infrastructure - extending sight without sharing in rule - then Chapters 8 and 9 explain why that constraint is necessary. They suggest that the central challenge is not how to make agentic systems govern well, but whether governance itself can remain visible and accountable once such systems are allowed to operate across the civic ceiling.

If that premise is mistaken, it should be challenged directly. But if it holds, then the path forward for Civic AI must begin not with greater capability, but with clearer limits.

For the constructive alternative, see:
"Civic Digital Visibility First: The Governing Principle for All Truly Civic AI."

- CHAPTER 8 -

THE DANGERS of AGENTIC AI in AMERICAN GOVERNANCE

I. The Temptation of Efficiency

If machines can drive cars, trade stocks, and write code, why shouldn't they help run government? For many Americans, that question does not feel theoretical. It feels like common sense born of frustration. They wait on hold with public offices. They navigate forms that seem designed to confuse. They watch crises unfold faster than agencies can respond and wonder why a country that can route millions of packages overnight struggles to process a permit in less than six weeks.

Under those conditions, the promise of agentic artificial intelligence in governance can sound almost merciful. What if digital "coworkers" could clear backlogs, route cases, and keep offices open 24 hours a day? What if proactive systems could fill out forms automatically, check eligibility in real time, and nudge people toward benefits they might otherwise miss? What if optimizers could scan thousands of projects and tell us, without politics, which roads, schools, or bridges should be fixed first?

It is important to take that hope seriously. Democracies do not win trust by romanticizing delay or defending every inefficiency as a virtue. In hospitals, courts, and welfare offices, slow decisions can mean real harm. In an age of tight budgets and rising expectations, public servants themselves often long for tools that could relieve them of repetitive work and let them focus on what truly requires human judgment.

Yet there is a reason earlier generations did not define a republic as "a government that does things quickly." Democracies are not designed to maximize efficiency. They are designed to limit and balance power. Procedures, debate, and even a certain amount of friction are not accidents in a constitutional order. They are how a free people keeps authority visible, contestable, and answerable for its use. Speed can be welcome. But when speed erases visibility, the cost is legitimacy. A free people does not merely ask how quickly power moves, but how clearly it can be seen.

What begins as assistance rarely remains so. Systems that help us decide have a quiet tendency, over time, to begin deciding for us. In the chapters that follow, this book will argue for a new class of Civic AI tools. They are non-agentic instruments of sight, not engines of decision. Before that argument can stand, however, we must confront an uncomfortable possibility. If we invite agentic systems into the machinery of government in the name of efficiency, they will not simply make old processes faster. Over time, they will change what it means to decide at all. That is the danger this chapter examines - not as a matter of misuse, but as a matter of structure.

II. Agentic Systems and the Civic Ceiling

Before we can assess the dangers of agentic AI in governance, we must be precise about what we mean by "agentic" and about the line this book treats as constitutionally non-negotiable. Vagueness is itself a risk here. If every automated spreadsheet is called "agentic" or every database query is called "Civic AI," then the terms lose the very clarity democracies need when new powers approach public authority.

In this work, a system counts as agentic when it does more than represent, retrieve, or illuminate information. It is agentic when it can, on its own initiative and within a defined scope, select actions, change states, or steer processes without fresh human judgment at each consequential step. An agentic system may open a case, close an application, route a file, trigger an alert, or assign a rank that other parts of government then treat as operative. Its defining feature is not intelligence, but delegated discretion. It is allowed to move.

By contrast, this book reserves the term Civic AI for a far narrower and more disciplined class of tools. Civic AI, as defined here, is structurally non-agentic. It may extend human perception, assemble and cross-reference records, detect patterns, and, when explicitly asked, provide analysis that clarifies options or consequences. It may clarify the consequences of choices, but it may never substitute for the act of choosing. Its vocation is illumination, not rule.

This distinction implies a civic ceiling - a boundary beyond which no machine may exercise practical discretion without eroding the conditions of self-government. Below that ceiling lie domains of function: logistics, formatting, internal scheduling, technical monitoring, and other forms of mechanical assistance that do not, in themselves, determine who is heard, who is served, or who is constrained. Above that ceiling lie domains of judgment: rights, eligibility, enforcement, durable access to public goods, and the structuring of speech, participation, and political choice.

The central claim of this chapter is simple. Systems that cross this civic ceiling - by deciding who qualifies, who is prioritized, who is watched, or how coercive power is applied - may be technologically impressive and administratively useful. They may be discussed as tools of public administration. But they cannot rightly be called Civic AI, nor can they be admitted near democratic authority without displacing it. Civic AI, by definition in this framework, occupies only one of three roles in a constitutional order: humans judge, bounded systems execute narrow functions below the civic ceiling, and Civic AI helps all parties see what is happening in time to decide. Anything more is not illumination; it is governance.

If this line feels severe, the rest of the chapter exists to test it. We will not ask whether agentic approaches can be built, or whether they might increase efficiency in some measure. We will ask whether, once admitted across the civic ceiling, they can avoid altering who effectively governs. Under that test, we will find that the danger is not occasional misuse, but structural drift.

III. The Structural Drift Principle

The risks of agentic systems in governance do not arise only from bad actors, biased data, or flawed implementation. They arise from a quieter and more stubborn pattern in how delegated systems evolve over time. This pattern can be stated simply:

- Systems that assist begin to recommend.
- Systems that recommend begin to be followed.
- Systems that are followed become embedded.
- Systems that become embedded begin to govern.

At the outset, an agentic tool may be introduced only to "help" human decision-makers manage complexity or volume. Its outputs are framed as suggestions, its role as advisory. But as time passes, two forces converge. First, institutions under pressure learn to rely on whatever reduces workload and controversy. Second, the very existence of a system that appears more consistent, more data-driven, or more objective makes it harder to depart from its guidance without justification. Deference becomes the norm, even when it is not formally required.

Once deference hardens into habit, the system's role changes. What began as one input among many becomes the practical baseline. Human judgment still exists, but it operates within a narrowed corridor defined by the machine's categories, rankings, or alerts. Over time, organizations are re-structured around these defaults: workflows, oversight routines, and even legal defenses begin to presume that "following the system" is the ordinary, prudent course. At that point, the tool is no longer merely assisting authority. It is helping to constitute it.

This book calls that progression structural drift. It does not depend on ill will. It is the expected behavior of delegated systems operating under institutional pressure. In private life, such drift may be unfortunate. Near democratic authority, it is disqualifying. Any agentic system allowed to operate across the civic ceiling will, if it proves even modestly useful, tend over time to move from support to substitution. That is why, in this framework, non-agentic design is not a preference. It is the constitutional precondition for anything that wishes to be called Civic AI.

IV. The Emerging Ecosystem of Agentic Governance Systems

The Structural Drift Principle would matter little if agentic systems near public authority were only speculative. They are not. Across the United States, a recognizable ecosystem of approaches is already being proposed, piloted, and in some cases deployed in and around government. They are described in hopeful language - GovAI, digital government, AI-enabled public services, agentic workflows - but structurally they share one feature: delegated discretion operating close to the civic ceiling.

For clarity, this chapter will group these efforts into six families. The labels are analytic, not official. They are offered to show that behind varied branding and technical detail, a small number of recurring patterns are emerging.

1. Automated Eligibility and Risk Systems:

The first family consists of systems that determine or heavily influence who qualifies, who is flagged, and who receives. These tools use past data to generate scores or classifications that guide decisions about individuals: eligibility for benefits, perceived risk in criminal justice or child welfare, likelihood of fraud in public programs, or priority in inspections and enforcement.

In formal terms, such systems transform data into discrete categories - approved or denied, low risk or high risk, routine or exceptional - that other institutional processes then treat as operative. Even when a human remains "in the loop," the default path is clear: the score is presumed correct unless overridden. Delegated discretion resides not only in explicit auto-approval or auto-denial, but in the way scores quietly reset the starting point of judgment. The system has redrawn, in practice, who the state sees as trustworthy, deserving, or dangerous.

2. Resource Prioritization and Planning Optimizers

A second family focuses less on individuals and more on where public resources go first. These are optimization and planning systems that rank projects, communities, or interventions according to specified objectives. They may seek to maximize cost-effectiveness, reduce measured disparities, improve resilience, or satisfy composite indices of "need" or "impact."

Here, the agentic element lies in the production of ordered lists or recommended portfolios that quickly become the default structure for budgets and plans. Once accepted as the rational way to allocate scarce resources, these rankings function as de facto policy. The choice of objective function - what to maximize, what to minimize, what to count and what to ignore - is itself a political judgment. Yet after the function is fixed, its outputs present as technical necessity rather than as the contested trade-offs they truly encode.

3. Agentic Workflows and Digital Coworkers

A third family operates inside agencies under the banner of workflow automation and "digital coworkers." These systems do not only store or summarize information. They route cases, open and close tasks, trigger follow-ups, escalate files, and in some instances move multi-step administrative processes forward with minimal human intervention. Their purpose is to clear backlogs, standardize handling, and ensure that no case "falls through the cracks."

In this domain, delegated discretion appears in timing and sequence. Who is contacted first, which applications are flagged as incomplete, which files are quietly closed after a threshold of non-response - all become emergent properties of routing logic and system rules. Over time, the practical experience of a right or benefit - how many hurdles must be cleared, how long a person must wait, how likely a case is to be revived - depends less on formal law than on how these internal agents manage queues and thresholds.

4. Autonomous Citizen-Service Agents

A fourth family sits directly at the interface between citizen and state: autonomous or semi-autonomous service agents. These systems go beyond answering questions or providing information. They fill out forms, check eligibility against existing records, submit applications, open cases, schedule appointments, and sometimes resolve service requests end-to-end. Some are explicitly designed to act "on behalf of" the resident.

Here, discretion is exercised in the interpretation of intent and in the construction of a path. The agent decides what the person meant to do, which programs or remedies are most relevant, which steps are necessary, and when the interaction is complete. Participation shifts from a citizen taking public actions in their own name to a system acting as their procedural surrogate. Misunderstandings, omissions, and biases in this surrogate's behavior may be hard to detect, especially for those with the least capacity to question the outcome.

5. Autonomous Security, Fraud, and Cyber Agents

A fifth family is justified in the language of protection: cybersecurity agents, fraud-detection engines, and anomaly-response systems. These tools monitor networks, transactions, and flows of activity continuously. They are designed not merely to log anomalies, but to act on them - to isolate machines, block accounts, trigger holds on payments, or open targeted reviews, often at machine speed.

In these settings, discretion is expressed through thresholds and automated responses. The system decides what counts as suspicious, when an alert warrants containment rather than observation, and how much collateral disruption is acceptable in the name of safety. Human review may occur after the fact, but the initial move - lock, freeze, flag, or sever - has already changed the practical situation of those affected. "Better safe than sorry" ceases to be an explicit human posture and becomes the standing behavior of a machine.

6. Advisory-to-Agentic Governance Tools

The final family begins life in the least threatening form: advisory and analytic systems that help policymakers or administrators make sense of complex evidence. These tools synthesize documents, model scenarios, compare options, and generate ranked lists of policies or actions. At first, they are framed as decision-support - another lens through which humans can view the field.

Over time, however, advisory systems tend to move along a familiar spectrum, following the same pattern described earlier: from assistance to recommendation, from recommendation to default, and from default to embedded governance. Recommendations become defaults. Defaults become automated in routine cases. Exceptions become rare. The tool that once merely described or suggested quietly shapes which options are even considered and what counts as a "reasonable" choice. In this way, design decisions about what is modeled, which values are encoded, and how options are ranked begin to steer the civic agenda itself.

7. Review of the Emerging Ecosystem of Agentic Governance Systems

Taken together, these six families constitute the practical field this chapter addresses. They differ in domain and rhetoric, but they share two structural traits: they operate close to or across the civic ceiling, and they embody the pattern of structural drift. They differ in function, but converge in consequence: the quiet relocation of judgment from human institutions to system behavior. Each promises assistance. Each is built to be useful. And, as the next section will argue, each contains within itself a characteristic way of moving from support to substitution that no democracy can safely ignore.

Taken on their own, these six families might still seem like a loose collection of innovations - each with its own promise, each with its own risks to be managed. What the Structural Drift Principle reveals is that they share a common trajectory. Each begins as assistance at or above the civic ceiling. Each embeds delegated discretion into decisions about who counts, who is served, who is watched, or which options are even considered. And each, if it proves useful, is pulled by ordinary institutional pressures toward becoming the place where practical judgment actually occurs.

Seen together, these six failure modes are not a scatter of unrelated problems. They are six variations on one theme: authority sliding, step by step, from visible human institutions into system behavior that is hard to see, harder to contest, and hardest of all to reverse once organizations have grown around it. That is structural drift in civic form. And because each family crosses or presses against the civic ceiling - exercising delegated discretion in domains of rights, eligibility, enforcement, access, and agenda-setting - each is, within this framework, constitutionally inadmissible as Civic AI, however effective or well-intentioned it may appear.

AGENTIC AI—FLOW OF FAILURE in AMERICAN GOVERNANCE

Across the United States, six families of agentic systems are emerging in and around governance. Each promises efficiency; each contains a characteristic democratic failure.

First, **automated eligibility and risk systems** promise faster, more consistent, and "objective" decisions about who qualifies, who is flagged, and who receives public benefits or scrutiny. They use past data to generate scores that other processes then treat as operative. The structural failure here is a hidden re-weighting of who counts: scores quietly redraw the civic boundary of trust and standing, shifting who is treated as deserving or dangerous from visible law and debate into opaque model behavior. In doing so, they breach equal civic standing and invite the moral scoring of persons.

Second, **resource prioritization and planning optimizers** promise "rational," data-driven allocation of limited funds and projects to maximize impact, efficiency, or fairness across communities. They generate ordered lists and recommended portfolios that quickly become the default structure for budgets and plans. Their structural failure is optimization without representation: political trade-offs are encoded in objective functions and then presented as technical necessity, so de facto policy is made inside algorithms rather than in legislatures. Human judgment over ends is displaced, and contested value choices are quietly outsourced.

Third, **agentic workflows and digital coworkers** promise reduced backlogs, standardized handling, and seamless case management through automated routing, escalation, and closure. They route cases, open and close tasks, and advance multi-step administrative processes with minimal human intervention. Their structural failure is procedural drift into invisible policy: the lived experience of rights and benefits - delays, hurdles, odds of revival - is quietly rewritten by workflow logic never authorized as law. Decisions are effectively made inside process code, eroding visibility of mandate and accountability.

Fourth, **autonomous citizen-service agents** promise frictionless, end-to-end access to services by allowing systems to act "on behalf of" residents - completing forms, checking eligibility, and submitting requests. Here the structural failure is substitution of voice: participation is mediated by a procedural surrogate that interprets intent and chooses paths, so direct civic agency is replaced by system behavior that is hard to see or contest, especially for the least powerful. Human agency in civic acts is displaced, and contestability is weakened.

Fifth, **autonomous security, fraud, and cyber agents** promise continuous protection of infrastructure and funds through real-time anomaly detection and automatic containment, holds, or targeted reviews. They monitor networks and flows of activity and act on anomalies at machine speed. Their structural failure is emergency logic without politics: locking, freezing, and flagging become standing machine behaviors rather than explicitly owned human choices, normalizing quasi-coercive actions without visible warrants of judgment and straining rights protections under stress.

Sixth, **advisory-to-agentic governance tools** promise better policymaking through summaries, simulations, and ranked options that make complex evidence more tractable and "evidence-based." They begin as decision-support systems but tend to move from assistance to recommendation, from recommendation to default, and from default to automation. Their structural failure is soft capture of the agenda: recommendations slide into defaults and defaults into embedded governance, so the real power lies in what is modeled and ranked, while elected officials govern within a pre-framed corridor of "reasonable" options. Agenda-setting is quietly outsourced, and pluralism narrows.

The following map summarizes that pattern. For each family, it sets out the efficiency claim that makes the system attractive, the named failure mode that emerges as drift takes hold, and the underlying civic principle strained or breached. All six operate where delegated discretion and democratic judgment meet. The point is not that every deployment will immediately collapse these safeguards, but that the structure of each approach makes the same kind of failure more likely the longer it remains in place.

AGENTIC AI - FLOW OF FAILURE in AMERICAN GOVERNANCE (Chart)

All six families operate at or across the civic ceiling, where delegated discretion begins to displace visible human judgment.

Agentic system family	Efficiency claim (the temptation)	Structural failure (democratic break)
Automated eligibility & risk systems	Faster, more consistent, and "objective" decisions about who qualifies, who is flagged, and who receives public benefits or scrutiny.	**Hidden re-weighting of who counts.** Scores quietly redraw the civic boundary of trust and standing, shifting who is treated as deserving or dangerous from visible law and debate into opaque model behavior. This breaches equal civic standing and invites moral scoring of persons (Gates 4, 27, 33).
Resource prioritization & planning optimizers	"Rational," data-driven allocation of limited funds and projects to maximize impact, efficiency, or fairness across communities.	**Optimization without representation.** Political trade-offs are encoded in objective functions and then presented as technical necessity, so de facto policy is made inside algorithms rather than in legislatures. This displaces human judgment over ends and outsources contested value choices (Gates 11, 14, 26).
Agentic workflows & digital coworkers	Reduced backlogs, standardized handling, and seamless case management through automated routing, escalation, and closure.	**Procedural drift into invisible policy.** The lived experience of rights and benefits—delays, hurdles, odds of revival—is quietly rewritten by workflow logic never authorized as law. This obscures where decisions are really made and erodes visibility of mandate and accountability (Gates 3, 15).
Autonomous citizen-service agents	Frictionless, end-to-end access to services by letting systems act "on behalf of" residents to complete forms, check eligibility, and submit requests.	**Substitution of voice.** Participation is mediated by a procedural surrogate that interprets intent and chooses paths, so direct civic agency is replaced by system behavior that is hard to see or contest, especially for the least powerful. This displaces human agency in civic acts and weakens contestability (Gates 5, 25, 42).
Autonomous security, fraud, and cyber agents	Continuous protection of infrastructure and funds through real-time anomaly detection and automatic containment, holds, or targeted reviews.	**Emergency logic without politics.** Locking, freezing, and flagging become standing machine behaviors rather than explicitly owned human choices, normalizing quasi-coercive actions without visible warrants of judgment. This violates the refusal of autonomous coercive action and strains rights protections under stress (Gates 24, 30).
Advisory-to-agentic governance tools	Better policymaking through summaries, simulations, and ranked options that make complex evidence tractable and "evidence-based."	**Soft capture of the agenda.** Recommendations slide into defaults, and defaults into automation; over time, the real power lies in what is modeled and ranked, while elected officials govern within a pre-framed corridor of "reasonable" options. This quietly outsources agenda-setting and narrows pluralism (Gates 26, 48).

V. Efficiency and Legitimacy

To this point, the analysis may seem to place efficiency and democracy in simple opposition, as if anything that speeds public work must be suspect. That is not the claim. In many settings, better tools, clearer information, and thoughtful automation can ease genuine burdens. A benefits office that reduces paperwork, a court that shortens needless delays, or a city that coordinates repairs more smoothly may all serve justice rather than weaken it. The six families described here are compelling in part because they address real institutional strain.

The difficulty is not that these systems fail to deliver gains. It is that they deliver those gains by moving discretion upward, across the civic ceiling, into opaque machinery that is no longer answerably human. Structural drift converts efficiency gains into a quiet re-location of judgment from visible institutions to embedded defaults in code and workflow. The more an agency depends on such systems for basic functioning, the harder it becomes in practice to refuse their outputs, contest their assumptions, or dismantle them without crippling essential services. What began as a choice to adopt a tool hardens into a dependency that effectively governs. At that point, efficiency is no longer simply a feature of governance; it has become a pathway to its quiet redefinition.

Democracies cannot accept that trade. Their first obligation is not to do everything faster, but to keep authority anchored where it can be seen, challenged, and owned. Legitimacy rests on the assurance that consequential decisions about rights, access, and coercive power are made by people who can be named, questioned, and, if necessary, replaced - not by systems whose inner workings few can explain and no one person is responsible for. When agentic designs cross the civic ceiling, they offer a bargain: fewer backlogs, smoother surfaces, quieter conflict, in exchange for a slow blurring of who actually decides. Under this framework, that bargain must be declined. Non-agentic Civic AI may help human institutions see more clearly and move more promptly. It may not, and must not, be replaced by systems that decide in their place.

On “Limited Agency,” “Bounded Autonomy,” and “Human-in-the-Loop”

Some will reply that “limited agency” is enough - that if systems act only within narrow bands, the danger is reduced to a manageable level. But structural drift does not depend on breadth of mandate. Even narrowly scoped systems, once embedded around recurring judgments about who qualifies, who is prioritized, or who is watched, begin to define the practical meaning of those judgments. The question is not how large the zone of agency is, but whether any zone of delegated judgment has been admitted across the civic ceiling at all. Once that line is crossed, the same progression from assistance to governance still applies.

Others will argue for “bounded autonomy” - agentic systems carefully constrained by rules, safeguards, and oversight. Yet from a constitutional perspective, this simply relocates the problem. If a system can open cases, close applications, rank persons, or trigger coercive steps without fresh human judgment in each instance, it is exercising authority in domains that a republic has no warrant to delegate to machines. The presence of guardrails may reduce harm; it does not change the fact that decisions once made under visible law and accountable officials are now being made inside code and control logic.

A third reassurance is that "human-in-the-loop is enough." In practice, however, structural drift turns many such humans into ratifiers rather than deciders. When institutional routines, legal defenses, and performance metrics all treat "following the system" as the prudent default, the burden of proof shifts onto anyone who would depart from its outputs. Over time, the loop reverses: it is the machine that effectively sets the baseline, and the human who must justify exceptions. Formal presence in the loop does not guarantee that judgment still resides with people; it often masks the fact that authority has already migrated into the system's defaults.

For that reason, the issue is not how agency is limited, bounded, or overseen, but whether it is admitted at all in domains where judgment must remain visibly human. In such domains, the question is one of admissibility, not calibration.

VI. Constitutional Verdict: Agentic AI as Power Near Democracy

By this point, the pattern is clear. Agentic systems that operate at or above the civic ceiling do not merely assist democratic governance. Over time, they reconfigure it. Automated eligibility and risk engines, optimization tools, workflow agents, citizen-service surrogates, autonomous security systems, and advisory-turned-agentic platforms each embody the same structural drift: from assistance to recommendation, from recommendation to default, from default to embedded governance. However different their technical details, they all move practical judgment away from visible human institutions and into system behavior that is difficult to see, contest, or reverse.

Within the framework of this book, that pattern is not a technical flaw to be tuned away. It is a constitutional failure. Civic AI, as defined here, begins with constraint rather than capability. It is non-agentic by design. It may illuminate conditions, assemble records, and analyze patterns on request. It may never decide who qualifies, who is prioritized, who is watched, how coercive power is applied, or which civic options count as "reasonable." Those are domains of judgment, and judgment in a republic must remain visibly and answerably human. Any system that crosses this line, however efficient or well-intentioned, stands outside the Civic AI category. It is simply powerful administrative AI operating near democracy.

Applied to the six families mapped in this chapter, the constitutional verdict follows directly. Automated eligibility and risk systems fail by redrawing civic standing through opaque scoring. Resource optimizers fail by encoding political trade-offs as technical functions. Agentic workflows fail by rewriting the lived content of rights through unvoted procedures. Citizen-service agents fail by substituting a procedural surrogate for the citizen's own civic voice. Autonomous security and fraud agents fail by normalizing machine-driven quasi-coercive action. Advisory-to-agentic tools fail by capturing the agenda within a pre-framed corridor of options. Each breaches, in its own way, the separation of illumination and authority that the civic ceiling is meant to enforce.

This does not mean that a democracy can never use automation or advanced computation in administration. It means that when such systems are agentic in domains of judgment, they may not be treated as Civic AI, and they may not be admitted near public authority under the standards proposed

here. They may be regulated as ordinary administrative tools, scrutinized as sources of risk, or rejected as incompatible with self-government. But they cannot share in the name or vocation of Civic AI, whose sole legitimate role is to extend sight without sharing in rule.

The work of the next chapter begins there. If agentic AI must be treated as power near democracy rather than as Civic AI within it, then a different kind of system is needed to watch the edge where that power operates. Non-agentic Civic AI, properly bounded, can serve as a lantern at that boundary - monitoring structural drift, illuminating where authority is quietly shifting, and buying back time for human institutions to deliberate. It cannot co-govern. It can help a republic keep its own promises in the presence of new and tempting tools.

An American republic, Benjamin Franklin once warned, would endure only "if you can keep it." The agentic systems described in this chapter do not announce themselves as threats to that charge; they arrive as helpers, optimizers, and digital coworkers. Yet their structure pulls authority away from the visible, answerable human judgment that keeping a republic requires. Civic AI, in this framework, is permitted only one role: to help a free people see what they are doing in time to decide for themselves. Anything more, however sophisticated, belongs on the other side of Franklin's warning - not as an instrument of keeping the republic, but as one more way it can quietly be lost.

- CHAPTER 9 -

WATCHING THE EDGE: CIVIC AI and THE MONITORING OF AGENTIC AI DRIFT

As artificial intelligence becomes more capable, it is unsurprising that governments will be tempted to use increasingly autonomous, "agentic" systems to manage public work. Routing vehicles, scheduling repairs, adjusting energy loads, forecasting demand - these are the kinds of functional tasks that lend themselves to automation. When such systems remain within the domain of logistics and engineering, they raise familiar questions of safety, reliability, and procurement. They do not, in themselves, constitute Civic AI.

The danger arises when tools built for function begin, often quietly, to exercise something closer to authority.

A routing system that merely optimizes bus schedules is one thing. A routing system that gradually withdraws service from certain neighborhoods - converting efficiency into exclusion without visible public decision - is another. In practice, many of the most consequential shifts in democratic life do not arrive as laws or doctrines. They arrive as "upgrades," "optimizations," and "recommendation engines" that reshape who can participate, who is heard, and who is left waiting at the curb.
If agentic systems are permitted anywhere near the machinery of government, even at its logistical edges, a republic must cultivate a new form of vigilance: the capacity to see when such systems begin to move inward. This is one of the central civic responsibilities of non-agentic Civic AI.

Civic AI does not co-govern. It does not tune, direct, or control. Its vocation is simpler and more constitutional: to watch the edge of authority.

Within this framework, Civic AI carries three linked responsibilities.

Monitoring Duty: Civic AI observes the aggregate behavior and outputs of agentic systems operating within or alongside public institutions. This observation is conducted strictly at the level of systems and patterns, never at the level of individuals. It tracks how such systems redistribute access, burden, and institutional dependence over time - for example, whether automated scheduling systems consistently reduce service to particular districts, or whether triage tools shift waiting-time burdens onto specific populations.

Boundary-Watch Duty: Civic AI detects when these patterns approach the line of civic judgment. That line is crossed whenever a system begins, in practice, to influence rights, eligibility, enforcement, speech conditions, or durable access to core public goods such as education, justice, health, or democratic participation. What appears as neutral optimization can, over time, become a de facto allocation of civic standing.

Alert-Only Duty: When such drift is detected, Civic AI performs one function only: illumination. It issues clear, non-binding notices to named human authorities and oversight bodies, explaining what has changed, why the shift may constitute movement toward civic authority, and which questions now require human deliberation. It does not recommend remedies, rank outcomes, or trigger automated responses. Machines may reveal that a boundary is being tested. Only people may decide how to respond.

This role must remain structurally bounded.

Civic AI may observe agentic systems, but it may not interact with, direct, or modify them. It must remain outside any operational control loop. The moment it begins to act upon the systems it observes, it ceases to be a Civic AI and becomes another agent within the system it was designed to illuminate. This constraint is not a limitation. It is the source of its legitimacy.

Agentic systems share a structural tendency: they do not remain static. Systems designed to assist begin to recommend. Systems that recommend begin to be followed. Systems that are followed become embedded. And systems that become embedded begin, often invisibly, to shape outcomes that were once the product of human judgment.

This progression is rarely declared. It emerges gradually - through convenience, efficiency, and institutional adaptation. Its effects, however, are cumulative. What begins as support can become substitution.

A republic cannot rely on intention alone to guard against this transformation. It must be able to see it. Civic AI provides that sight.

If agentic systems are to exist at the margins of governance, then a republic worthy of the name must also build instruments capable of seeing when those systems begin to move inward. Without such visibility, boundaries erode silently, authority shifts without acknowledgment, and democratic accountability weakens without clear cause.

With visibility, the boundary remains real.

The principle is simple: Delegation may be acceptable in domains of function. It is not acceptable in domains of judgment.

Within democratic governance, every consequential decision about who counts, who is heard, and who belongs must remain visibly and answerably human.

Civic AI helps ensure that this line remains clear - not by holding power itself, but by making power's movements visible in time to be governed.

It watches the edge.

Author's Related Work: This essay draws on Christopher Hunt Robertson's books-in-progress, *Civic AI Visibility Infrastructure for U.S. Democracy* (Volumes I–III). The project advances a "Civic Digital Visibility First" architectural stack: (1) Six Rings of Digital Invisibility and Civic Destabilization (problem diagnosis); (2) the Ben Franklin Civic Data Observatory (border-facing national visibility infrastructure); (3) the Fifty-Gates Democratic Permissions Threshold (constitutional admissibility for all Civic AI); (4) non-advisory Civic AI Archivists and their Democratic Civic Impact Assessments (illumination-only analytical reports); and (5) Civic AI Micro Moral Alignment (a formative discipline for institutional AI). *** These proposals address civic digital invisibility, not general data-privacy regulation.